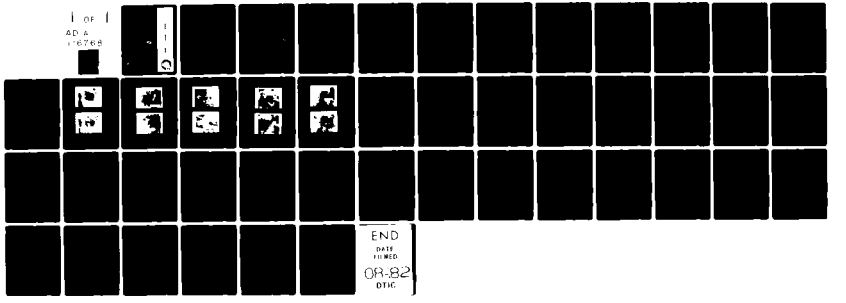
AD-A116 768 ARMY ENGINEER TOPOGRAPHIC LABS FORT BELVOIR VA F/G 14/5
AN ANALYSIS OF THE MAX-MIN TEXTURE MEASURE.(U)
JAN 82 M A CROMBIE, R S RAND, N FRIEND
UNCLASSIFIED ETL-0280 NL



ETL–0280

An analysis of the max–min texture measure

Michael A. Crombie
Robert S. Rand
Nancy Friend

JANUARY 1982

U.S. ARMY CORPS OF ENGINEERS
ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA 22060

E T L

U.S. ARMY ENGINEER TOPOGRAPHIC LABORATORIES
ETL

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

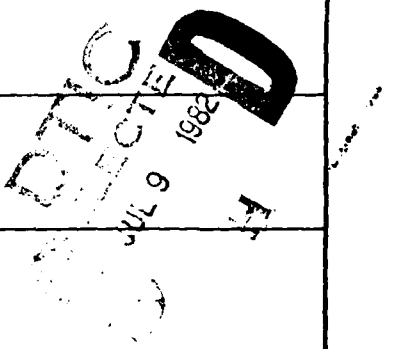
UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ETL-0280	2. GOVT ACCESSION NO. AD-A116 768	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) AN ANALYSIS OF THE MAX-MIN TEXTURE MEASURE		5. TYPE OF REPORT & PERIOD COVERED Research Note 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Michael A. Crombie Robert S. Rand Nancy Friend		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, Virginia 22060		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 4A762707A855-B-0026
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories Fort Belvoir, Virginia 22060		12. REPORT DATE January 1982 13. NUMBER OF PAGES 42
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)



18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

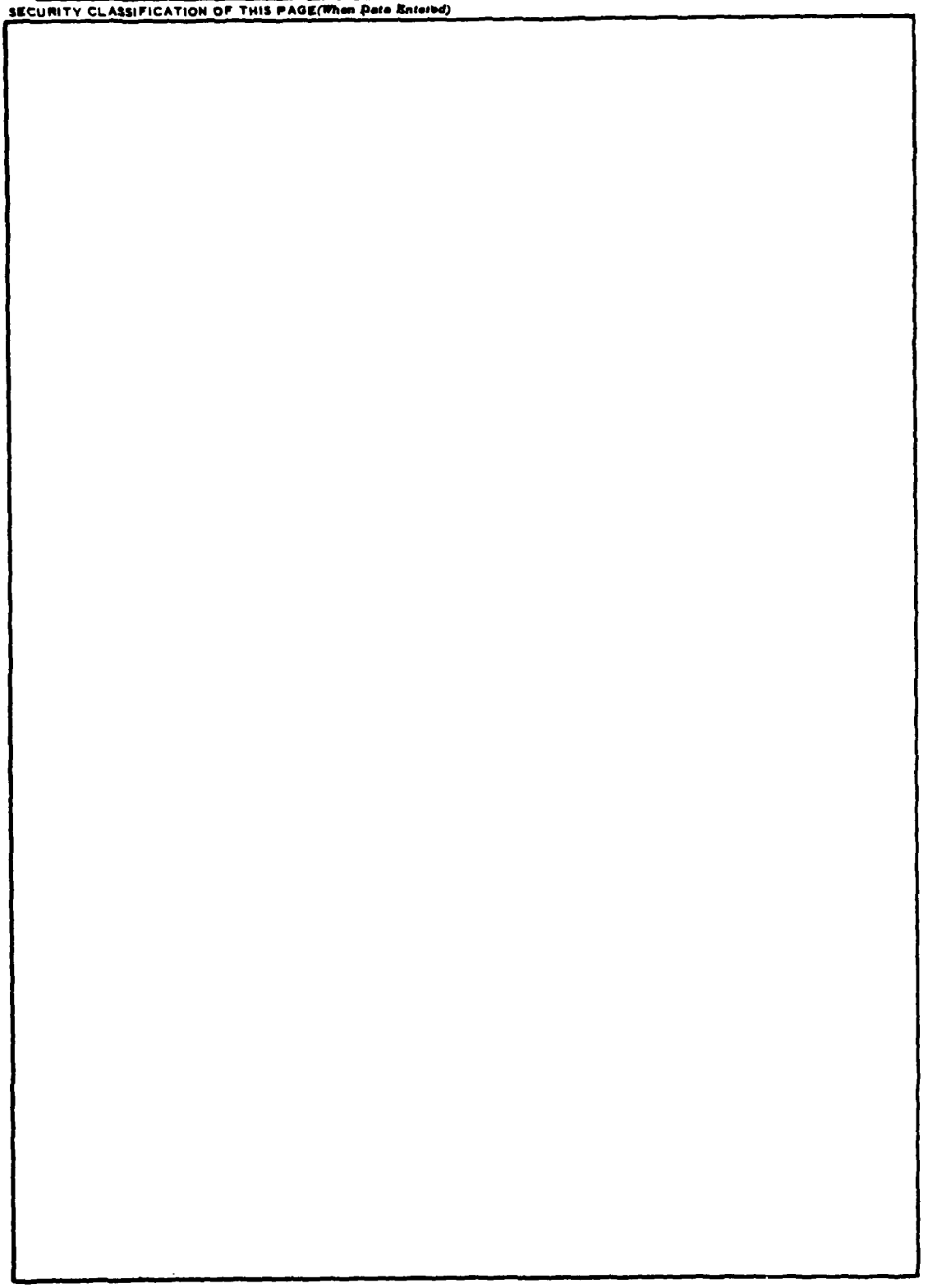
Divergence
Image Texture
Maximum Likelihood
Principal Component
Relaxation Labeling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The Max-Min texture measure was tested in an auto evaluation experiment and found to be an excellent signature for discriminating among several natural features and for isolating cultural detail from natural detail. This is especially true if the classification results are refined by a relaxation labeling exercise.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)



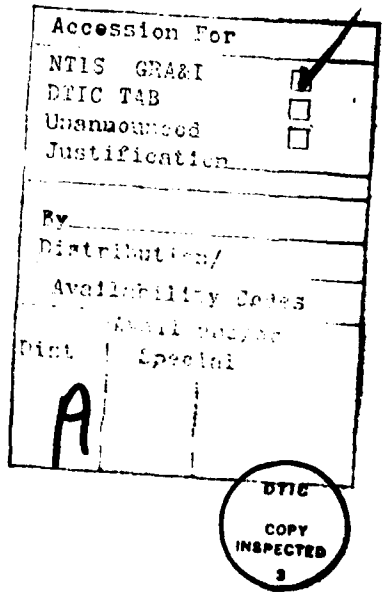
PREFACE

This study was conducted under DA Project 4A762707A855, Task B, Work Unit 0026, "Topographic Mapping Techniques."

The study was done during 1980 under the supervision of Mr. Dale E. Howell, Chief, Information Science Division; and Mr. L. A. Gambino, Director, Computer Sciences Laboratory.

Special thanks are extended to Mr. Thomas A. Hay who assisted with the computer programing.

COL Edward K. Wintz, CE was Commander and Director, and Mr. Robert P. Macchia was Technical Director of the Engineer Topographic Laboratories during the report preparation.



CONTENTS

ILLUSTRATIONS

TABLES

AN ANALYSIS OF THE MAX-MIN TEXTURE MEASURE

INTRODUCTION

In lieu of a satisfactory mathematical description of image scene content, the U.S. Army Engineer Topographic Laboratories (ETL), Computer Sciences Laboratory (CSL) has embarked on a heuristic approach to information extraction from digital and digitized images, in response to a Defense Mapping Agency (DMA) requirement for improving current methods of determining Digital Landmass System (DLMS) data. This research note describes the first of many experiments in which texture is used to specify a signature that, in turn, is used to discriminate among a variety of candidates for classification.

The scene recognition problem under study pertains only to mapping, charting, and geodotic data, of which the DLMS is a subset. The initial investigations are concerned primarily with passive sensor records rather than with active ones. Consideration is given to limitations imposed on any process by (1) an inadequate mathematical model that causes false starts and *ad hoc* solutions; (2) by the computer, which allows a local-only view of the scene rather than a global one; and (3) by the sensor record itself, which is a two-dimensional distorted version of a three-dimensional world wherein detail is obscured and, in many cases, not even resolved.

The CSL approach toward scene recognition is an attempt to combine the heretofore separate processes of elevation determination and scene classification into a cooperative venture that uses one process to strengthen the other and enables manual intervention through computer display devices. The approach has evolved over the last several years as a result of in-house work and contractor studies at ETL. An overview of the approach was first expressed in a paper presented to the Congress of the International Federation of Surveyors,[1] and later in a paper presented to the Naval Postgraduate School.[2]

---

[1]Michael A. Crombie and Lawrence A. Gambino, "Digital Stereo Photogrammetry," Prepared for Congress of the International Federation of Surveyors (FIG), Commission V, Stockholm, Sweden, June 1977.

[2]Lawrence A. Gambino and Michael A. Crombie, "Manipulation and Display of Digital Topographic Data," Prepared for the Second Symposium on Automation Technology In Engineering Drawing, Monterrey, Calif., November 1979.

One specific texture measure, namely the Max-Min measure, is evaluated in this research note. A modification of the conventional Maximum Likelihood Classifier, wherein multivariate normal populations are assumed, is used to classify the texture signatures. The modified version produces the R most likely candidates and R probabilities associated with the estimates rather than just the most likely candidate. The R probabilities are introduced to a postprocessing relaxation scheme in an attempt to reduce misclassifications and false alarms.

MATHEMATICAL DESCRIPTION OF THE EXPERIMENT

Everyone seems to know what texture is as long as a precise definition is not required. The hesitancy in producing a precise definition increases whenever the intuitive notions of texture must be modified to describe image texture. A variety of definitions and texture descriptors are available to choose from.[3] In the work at CSL, image texture has been regarded as a point and/or line pattern that is somewhat repetitive. The minute pattern of detail can be described by measures of spatial image structure that convey the idea of varying degrees of coarseness. If texture is regarded as the relative frequency of local extremes in pixel intensity, then one measure is the Max-Min texture measure, which is the number of local gray level maximums and minimums along a one-dimensional scan.[4]

Max-Min Texture Measure • Consider a one-dimension array of quantized density values (gray shades). Density is used rather than transmission, because a uniform multiplicative change in light intensity will not affect a density gray shade change over a uniform texture. For example, if one part of a field is shaded and another fully illuminated by the sun, then Max-Min texture measures extracted from the two should be equal. Unlike LANDSAT, which provides four spectral gray shades for each pixel, texture measures associated with a pixel from a panchromatic image must be estimated from data at and around the pixel. The method of determining the Max-Min texture measure is described next.

---

[3]Robert M. Haralick, "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, Vol. 67, No. 5, May 1979.

[4]Owen R. Mitchell, Charles R. Myers, and William Boyne, "A Max-Min Measure for Image Texture Analysis, *IEEE Trans. Comput.*, Vol. C-25, April 1977.

A rectangular window of gray shades centered over the point in question is specified along with a fixed set of threshold values. Four types of scan lines are considered over the window, namely the set of horizontal lines, the set of vertical lines, the set of left diagonal lines, and the set of right diagonal lines. Essentially, a line of gray shades from any one of the four sets is selected along with one of the NT threshold values T. A smoothing process, which is a function of the specific T, is applied to the data in order to remove small amplitude changes so that only the significant extremes are retained and counted. The extremes that exceed T are counted and summed over the specific set of lines. Each of the four sets of lines will produce NT extreme counts. The next step in the process of generating Max-Min texture vectors is to ratio the number of extremes at one value of T to the number of extremes at the next larger value, thus producing four (NT−1) component vectors where each component is a nonnegative number less then, or equal to, one. The reason for this step is to produce texture data that is invariant to the absolute number of extremes. There should then be a window size of dimension NW that produces consistent texture data in which an increase in dimension does not improve consistency. This is an important consideration, since small windows produce better output resolution and require less computing time than larger windows.

Maximum Likelihood Classification Rule • The maximum likelihood rule as described in appendix B was used in this experiment for the following reasons. The Baysian classifier (see appendix A) from which it was derived has a logical and intuitive appeal. The specification of the multivariate normal population in the Baysian rule rested on familiarity with the process as well as a certain comfort derived from the Central Limit Theorem. Finally, the procedure is well documented, and it lends itself well to the modifications employed in this experiment.

The Max-Min algorithm described above produces four texture vectors, one for each linear direction within a window. There are various ways the data could be processed. For example, one direction could be chosen and used throughout. This would reduce the compute time considerably. Also, the four texture vectors could be averaged and the average vector characterized. This procedure would not be as effective in the reduction of computing time, but areal texture would be measured. Finally, the four texture vectors could be processed separately and the final characterization derived from the four results. This procedure is the most time consuming, but the algorithm is simple enough so that if it were proven worthwhile, a hardwired device could be developed. The last approach was selected and is described next.

Consider any one of the four texture vectors. Instead of choosing the most likely class as described in appendix B, choose the R most likely classes. This was done in the following manner:

Produce the N decision values for each of the four directions. Normalize the vectors of decision values, that is let

$$\bar{f}_s = D'_{1s}(\bar{X}), D'_{2s}(X), \ldots, D'_{Ns}(\bar{X})$$

S = 1, 4 directions

N: Possible number of classes

The $D'_{is}(\bar{X})$ are normalized versions of the decision functions described in appendix B.

$$D'_{is}(\bar{X}) = \frac{D_{is}(\bar{X})}{\sum_{i=1}^{N} D_{is}(\bar{X})}$$

Let $\bar{f} = \sum_{s=1}^{4} \bar{f}_s$ and choose the R largest of the N components of $\bar{f}$.

Normalize these data and produce the following output.

$$\left[\left(C_{i1},\ P(C_{i1})\right),\ \left(C_{i2},\ P(C_{i2})\right),\ \ldots \left(C_{iR},\ P(C_{iR})\right)\right]$$

C_{ir}: the rth most likely of the i = 1,N possibilities for $\bar{X}$.

The weights $P(C_{ir})$ will be regarded as probabilities in the relaxation scheme described next.

Relaxation Labeling • The relaxation scheme used in this experiment is an application of a nonlinear probabilistic model to relaxation labeling.[5] Consider a set of results from the classification process described. The output for each of the processed pixels is an R-tuple in which each component is made up of a label C_{ir} and an associated weight $P(C_{ir})$, where $i = 1,N$ possible classifications (labels) and $r = 1,R$ most likely labels for the specific point. Since each of the weights is greater than zero and since $\sum_{r=1}^{R} P(C_{ir}) = 1$, the weights can be regarded as probabilites.

The vector of probabilities is processed in parallel by reviewing each component with respect to its neighbors. Modifications to the probabilities are made to reflect neighboring pixel information and to reflect user-imposed weights and constraints. The constraints are realized by a user-defined relational matrix that describes the compatibility of neighboring classes. The probability update is expressed in the following formula:

$$P_l^{k+1}(C_{ir}) = \frac{P_l^k(C_{ir}) \left[1 + q_l^k(C_{ir})\right]}{\sum_{\xi=i1}^{iR} P_l^k(C_\xi) \left[1 + q_l^k(C_\xi)\right]}$$

$$q_l^k(C_\xi) = \sum_{g=1}^{G} W_g \sum_{\gamma=1}^{iR} a(C_\xi, C_\gamma) P^k(C_\gamma)$$

ξ = i1, iR: number of labels (classes)

k = 1, K: iterations of the process

l = 1, P: pixels to process

g = 1, G: neighbors used in the update

W_g: user assigned weights. Note that $\sum W_g = 1$.

$a(C_\xi, C_\gamma)$: NxN compatibility matrix.

[5] Azriel Rosenfeld, Robert A. Hummel and Steven W. Zucker, "Scene Labeling by Relaxation Operation," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-6, No. 6, June 1976.

The determination of the weights (W_g) and the compabibility values in $a(C_\xi, C_\gamma)$ are left up to the user. These inputs allow the user some control of the iterative update. For example, if $\omega_1 = 0$ (weight of pixel under modification), then the neighbors completely determine the classification. The compatibility factors are numerical values between minus one and plus one. A value of plus one for a (C_ξ, C_γ) implies that classes C_ξ and C_γ can occur together with no problem; whereas, a value of minus one implies that the existence of C_ξ at location 1 denies the existence of C_γ at neighboring pixels. A value of zero implies that the existence of label C_ξ at location 1 has no bearing on the existence of C_γ at neighboring pixels.

The intention here is to use the relaxation labeling concept as a means to provide the computer with a larger view of the scene by relating derived data in a rational manner. There are various ways that the algorithm can be modified. It is expected that modification will come about as practical experience is obtained. There are many aspects of the algorithm as defined above that need to be tested. For example, using $(1 + q)^\alpha$ with $\alpha > 1$ will speed up convergence; however, the effect of distant points will not be felt if the number of iterations is reduced. It should be noted that the compatibility matrix need not, and probably should not, be symmetric. It should also be noted that if the initial probability of a class is zero, then the algorithms as employed in this experiment will never revise the probability, even if every neighbor of the point insists on that classification.

NUMERICAL EXPERIMENTS

Text Regions • Ten digital subscenes (1024^2 x 8 bits) were extracted from the Digital Image Analysis Laboratory (DIAL) library for the texture analysis. The parameters of the taking geometry and of the scan procedure are described in a previous research note.[6] The set of 10 subscenes is composed of a near infrared exposure (IR) and a corresponding panchromatic exposure (PANC) for each of 5 scenes. The 5 scene pairs are shown in figures 1 through 5.

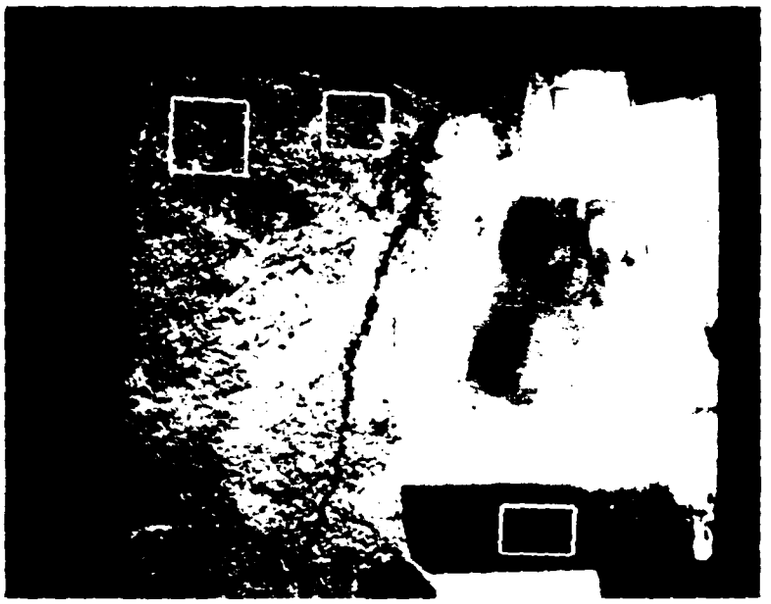
---

[6]Michael A. Crombie, *An Evaluation of Conventional Correlation Methods When Matching Infrared Imagery to Panchromatic Imagery*, U.S. Army Engineer Topographic Laboratories, Fort Belvoir, VA ETL-0195, August 1979, AD-A076 111.

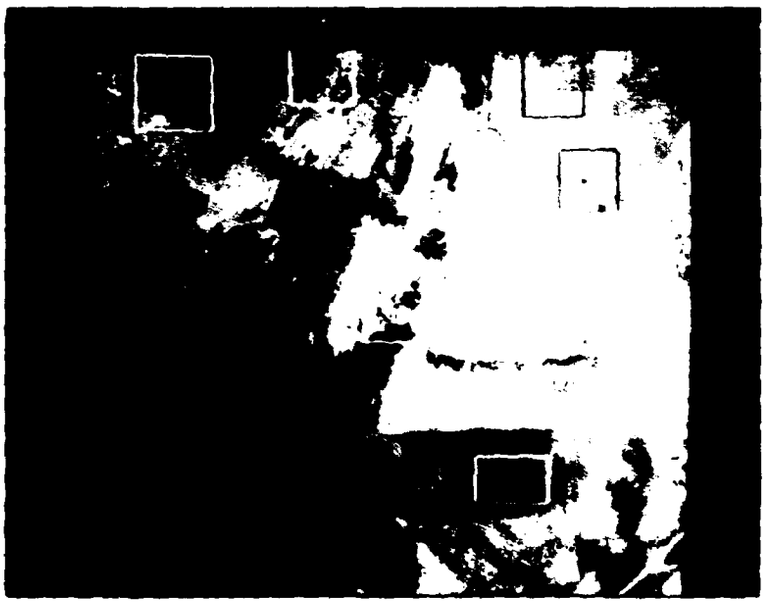
INFRARED

PANCHROMATIC

FIGURE 1. Scene A From Exposure 54.



INFRARED



PANCHROMATIC

FIGURE 2. Scene B From Exposure 54.

INFRARED

PANCHROMATIC

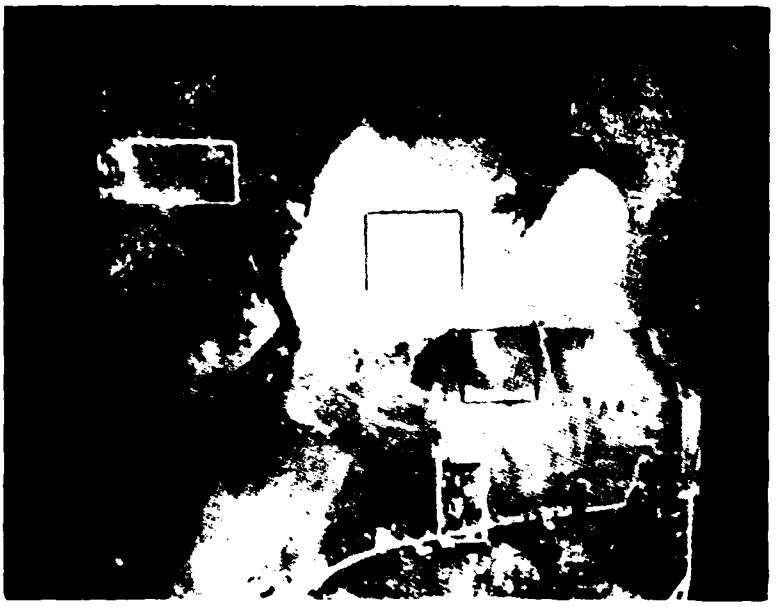
FIGURE 3. Scene C From Exposure 54.

INFRARED

PANCHROMATIC

FIGURE 4. Scene E From Exposure 54.

INFRARED



PANCHROMATIC

FIGURE 5. Scene H From Exposure 54.

The portions of the subscenes selected for the texture analysis were designated by rectangles. The scene content in each rectangular window was determined by looking at the IR scene on the DIAL display. Results of the visual designations are presented below. A regular pattern of points (every 10 lines and every 10 pixels) over each rectangular area was used to develop the texture signatures.

SCENE A

CLASS	TYPE
1	Building and Road
2	Gray Field
3	Rough Field
4	Heavy Forest
5	Light Field
6	Light Forest

SCENE B

CLASS	TYPE
1	Heavy Forest
2	Scrub
3	Field, Building, and Road
4	Dark Field
5	Light Field
6	Light Forest

SCENE C

CLASS	TYPE
1	Heavy Forest (Light)
2	Heavy Forest (Dark)
3	Light Forest
4	Light Field
5	Gray Field

SCENE E

CLASS	TYPE
1	Dark Field
2	Light Field
3	Heavy Forest
4	Scrub
5	Building and Road
6	Light Forest

SCENE H

CLASS	TYPE
1	Dark Field
2	Light Field
3	Scrub
4	Building and Road

Three window sizes were used to develop the texture signatures.

WINDOW	*GROUND FOOTPRINT (FEET)*
9 x 9	26 x 26
15 x 15	46 x 46
21 x 21	66 x 66

The 15 threshold values used to develop the Max-Min features were: 1, 3, 5, 7, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36 and 39.

The Max-Min texture feature vectors (14 components) were calculated and stored on disc for a variety of analyses.

Numerical Results • Four kinds of numerical results were derived from the texture features; namely component compression, divergence measures, confusion matrices before relaxation, and confusion matrices after relaxation.

Component Compression • A principal component analysis of the data was performed to determine whether a significant amount of information contained in the 14-component Max-Min texture feature vectors could be explained by fewer components. Results from the 648 separate component analyses are summarized in table 1.

TABLE 1. Principal Component Results by Scene.

		Sampling Window Size								
		9			15			21		
		PRINCIPLE COMPONENTS								
SCENE		3	5	7	3	5	7	3	5	7
A										
	PANC	72.5	83.9	91.2	75.8	88.0	94.2	77.6	89.8	95.3
	IR	73.1	83.8	91.2	75.2	87.2	93.8	75.9	88.0	94.1
B										
	PANC	73.9	84.2	91.2	75.0	87.0	93.4	75.7	87.4	93.4
	IR	75.5	85.2	91.7	74.0	85.7	92.2	74.4	85.5	92.0
C										
	PANC	68.9	81.8	90.1	64.8	79.7	88.1	63.3	78.8	87.7
	IR	70.3	82.3	90.3	66.7	81.3	89.6	67.9	82.1	89.7
E										
	PANC	73.8	84.6	91.6	74.7	86.7	93.3	71.4	85.5	92.8
	IR	73.3	84.4	91.5	67.4	80.9	89.1	65.4	80.2	88.7
H										
	PANC	70.7	82.2	89.7	75.1	87.4	93.8	78.2	90.6	95.5
	IR	71.4	82.8	90.4	74.7	86.1	92.9	77.1	89.2	94.8

The tabular entries are percentages of variation explained by the first 3, 5, and 7 principal components. The results are organized by scenes, by the two kinds of exposures, and by the sampling window size. The results from table 1 were averaged over scene and presented in table 2. The results of table 2 show little variation over window size or exposure type.

TABLE 2. Principal Component Results.

	9			15			21		
	3	5	7	3	5	7	3	5	7
PANC	72	83	91	73	86	93	73	86	93
IR	73	84	91	72	84	92	72	85	92

Divergence Measure • The amount of information, $I(C_I : C_J)$, contained in the N-dimensional vector $\bar{X}$ for discriminating in favor of class I over class J is described in appendix C. Similarly, the amount of information contained in $\bar{X}$ for discriminating in favor of class J over class I is $I(C_J : C_I)$, and by definition, $J(C_J, C_J) = I(C_I : C_J) + I(C_J : C_I)$ is a measure of the difficulty of discriminating between class I and class J. $J(C_I, C_J)$ is a nonnegative number and is called divergence.

Large values of $J(C_I, C_J)$ are indicative of strong discriminatory power, whereas small values of $J(C_I, C_J)$ indicate poor discriminatory power. The purpose of the divergence analysis was to determine if a worthwhile relation between classification error and the associated divergence could be developed so that a variety of texture parameters could be evaluated without performing the computing-intensive, maximum-likelihood evaluation and subsequent review of the resultant confusion matrices.

Two measures of association were calculated to determine whether a consistent relation between divergence and classification errors exists. The first was the linear correlation coefficient R_{DE}, which is a measure of the linear dependence between divergence and classification error. The second was Spearman's rank correlation coefficient P_{DE}, which is a measure of the degree of correlation between rankings. This statistic is calculated from the difference in rankings. The classification errors are ranked from 1 to M (smallest to largest). The divergence values are also ranked from 1 to M (largest to smallest). In both cases M is the number of combinations of the N possible classes taken two at a time. A summary of the results is presented in table 3.

TABLE 3. Classification Error Versus Divergence Statistics.

		R_{DE}			P_{DE}		
		SAMPLING WINDOW SIZE					
		9	15	21	9	15	21
SCENE							
A							
	PANC	−0.41	−0.50	−0.53	0.39	0.51	0.70
	IR	−0.37	−0.33	−0.52	0.48	0.45	0.61
B							
	PANC	−0.65	−0.58	−0.68	0.71	0.63	0.71
	IR	−0.12	−0.79	−0.63	0.23	0.74	0.64
C							
	PANC	−0.62	−0.73	−0.62	0.72	0.92	0.96
	IR	−0.69	−0.77	−0.81	0.82	0.87	0.94
E							
	PANC	−0.61	−0.78	−0.59	0.58	0.88	0.96
	IR	−0.45	−0.54	−0.35	0.49	0.88	0.95
H							
	PANC	−0.65	−0.92	−0.81	0.66	0.94	0.54
	IR	−0.50	−0.46	−0.59	0.43	0.26	0.26

Confusion Matrices Before Relaxation • The required statistical entities needed to evaluate the decision functions described in the section on the Maximum Likelihood Rule were computed from the stored Max-Min texture features. The derived population parameter estimates were used to classify the stored Max-Min texture features by designating the most likely of the R = 3 most likely classifications as the correct classification. The designated classifications were compared to the known classifications in order to develop a matrix of hits and misses. These results were modified to produce the confusion matrices given in appendix D. The 10 sets of confusion matrices are organized first by scene and then by exposure type. Each of the 10 tables of data is composed of 6 subtables organized by sampling window size and by maximum likelihood results and by relaxation results.

Confusion Matrices After Relaxation • The relaxation process described in the Relaxation Labeling section was applied to the derived vectors of most likely classes and associated probabilities. The following values were used to run the algorithms:

$$R = 3 \text{ most likely classes}$$

$$K = 5 \text{ iterations}$$

$$G = 25 \text{ neighbors used in the update}$$

$$a(C\xi, C\gamma) = \begin{cases} 1 \text{ if } \xi = \gamma \\ 0 \text{ if } \xi \neq \gamma \end{cases}$$

A 5 x 5 window centered over the point to be updated defined the G = 25 neighbors. In this example, the point to be modified was used in the update. The vector of probabilities associated with the center point was given a weight of W = 1/3. Each of the vectors of probabilities associated with the eight nearest neighbors was given a weight of W = 1/24. Each of the vectors of probabilities associated with the 16 next nearest neighbors was given a weight of W = 1/48. The weighting procedure was modified to handle the situation when points along the boundary of derived data (or points adjacent to the boundary) were updated. The results of the relaxation are presented in appendix D. The organization of the results is identical to those described in the previous section.

DISCUSSION

The training areas outlined on the scenes (see figures 1 to 5) were defined on the IR imagery by mathematicians, not experienced photo interpreters. For example, heavy forests were distinguished from light forests, and these from scrub, purely from appearance. A subjective evaluation of the density of trees was the criterion used. The same was true when for Scene C, heavy forest (light) and heavy forest (dark) were characterized and distinguished from light forest. Dark and light pertained to tone in the heavy forest descriptions. It was noted that the two appeared more dissimilar on the IR than on the PANC, yet the classification scores were similar, especially when Max-Min texture

features were extracted using 21 x 21 windows. It was also noted that the distinction between a light field and a dark field identified on the IR was less pronounced on the PANC. Again, the classification performances for the two exposure types were comparable, which indicates a difference in Max-Min texture between light and dark fields. It was noted after the experiment was well underway that the training region around the buildings and road (class 3) on Scene B was too large. The rectangle included too large an area of light and dark fields. The resultant poor classification of this area in Scene B is noticeable, especially when compared to results for small window sizes from the other scenes.

The purpose of the initial experiment in texture analysis was to evaluate the Max-Min texture signature as a means to sort out broad areas, such as forests and fields, and to isolate cultural detail from surrounding natural detail. It is not expected that texture will provide enough information to distinguish among the various DLMS cultural features. It does appear from the results to date that texture can isolate structures for subsequent identification by a photo interpreter on a display device. This is especially true if the preliminary results from the classification exercise are refined by relaxation. In fact, the relaxation process as shown in the confusion matrices presented in appendix D considerably reduced misclassification and false alarms.

The results to date are not extensive over scene types, scale, or exposure type. Five scenes were tested (actually subsets of those scenes) and those scenes were generally alike. Two exposure types (near infrared and panchromatic) with near-identical scales (1:70,000) were used. Three different window sizes were evaluated. The quality of the classification results and the quality of the relaxation results noticeably improved as the window size increased. It should be noted that the computing time lengthens as the window size increases and that the resolution of boundaries between, say, fields and forests will diminish as the window size increases. Owing to the lengthy computer runs associated with this experiment, only one set of thresholds was tested, namely that given in the section on test regions.

Two statistical studies were performed during the course of the experiment. The first was a compression study wherein the method of principal components was used to explain the percentage of variation accounted for by the first three, by the first five, and by the first seven principal components. Results averaged over scene, over exposure type, and over window size show that 72 percent of the variation is explained by the first three components, 85 percent by the

first five components and 92 percent by the first seven components. It should be noted that these results pertain to individual subscene statistics rather than to overall scene statistics. Classification results were not evaluated using the reduced texture signatures owing to the lack of time. Tests using compressed texture vectors should be conducted on subsequent experiments, since it is impossible to estimate results using the percentage of explained variation as a predictor.

An unsuccessful attempt was made to use the statistical measure of divergence as a measure of effectiveness for comparing one set of texture parameters to another. The correlation values R_{DP} presented in table 3 indicate a negative correlation (as divergence increases, classification error decreases); however, the values do not indicate a strong linear trend. Several sets of data were plotted to determine if a functional relation other than a linear one existed. Such a relation was not apparent from the plots. The results were ranked as described in the section on Divergence Measure to determine whether there was a consistency between the two sets of results. Only those values underlined in table 3 refute (at the 1 percent confidence level) the hypothesis that the two sets of rankings were random. Generally, there is some support for a relationship between divergence and classification error, but not enough to warrant doing away with the classification tests.

The Max-Min texture measure was tested and found to be an excellent signature for scene classification, especially when relaxation labeling processes are used to refine the classification results. The necessity of using regions about the point in question to estimate texture, rather than estimating texture from several pixel component values, as is done in LANDSAT, causes a loss of resolution in determining textural boundaries. The loss in resolution becomes more pronounced as the sampling window increases in size. The results of the tests to date indicate that a 9 x 9 window is too small for the problem at hand. The smaller windows, although more precise for determining textural boundaries, tend to measure the texture of the trees rather than texture of the forest.

The loss in resolution at textural boundaries caused by larger sampling windows can be recovered by employing three algorithms in a cooperative mode: (1) digital stereo compilation, (2) edge detection (thresholded gradient method), and (3) texture classification. In practice, stereo images exist that can be used to extract two estimates of texture for each point. The procedure is feasible if the process is executed along with the stereo compilation exercise. In the same manner, the edge algorithm can be used to produce two corresponding sets of

edge data. The processes can be integrated and controlled through known exterior data relating the stereo imagery and by relational statements through combined relaxational processes. For example, local slope information derived from X-parallax data can be used to modulate the classification process and *vice versa.*

CONCLUSIONS

1. *Max-Min is a practical measure* of image texture that needs further investigation.

2. Relaxation labeling appears to be a valuable method for removing noise and ambiguities from derived classifications.

3. Edge detection methods should be used, along with classifications derived from texture, to define boundaries between broad areas.

4. Texture can be used to isolate cultural detail from natural detail.

5. Divergence does not appear to be a worthwhile predictor for precise classification performance.

6. Relaxation methods should be extended to encompass edge enhancement, elevation refinement, and texture classifications in a cooperative mode.

APPENDIX A.

Bayes Classifier • Regard the classification process as part of a two-person, zero-sum game $G = (U, V, L)$, where U is an $(N \times 1)$ vector of strategies for the first player, V is an $(N \times 1)$ vector of strategies for the second player, and L is an $(N \times N)$ loss function. If the first player (Mother Nature) chooses a class $c_i \in U$ and if the second player (photo interpreter) choose a classification $c_j \in V$, then the second player loses l_{ij}. Suppose the first player selects class c_i based on the prior probability $p(c_i)$ and produces an $(M \times 1)$ Max-Min texture vector $\bar{X}$. The posterior probability that $\bar{X}$ belongs to class c_i is $p(c_i/\bar{X})$. If the unfortunate second player decides that $\bar{X}$ comes from c_j, he loses l_{ij}. From the second player's point of view, the pattern vector $\bar{X}$ could have come from any one of the N possible classes. His expected loss for assigning the signature to the j^{th} class is

$$r_j(\bar{X}) = \sum_{i=1}^{N} l_{ij}\, p(c_i/\bar{X}).$$

Suppose the second player (numerical classification process acting in place of the photo interpreter) calculates $r_j(\bar{X})$; $j = 1, N$ and assigns the pattern $\bar{X}$ to the class with the smallest $r_j(\bar{X})$. If this is done for all $\bar{X}$, then the total expected loss with respect to all decisions will be minimized. Such a classifier is called a Bayes Classifier.

The Bayes formula for posterior probability is

$$p(c_i/\bar{X}) = \frac{p(c_i)\; p(\bar{X}/c_i)}{p(\bar{X})}$$

where

$$p(\bar{X}) = \sum_{k=1}^{N} p(c_k)\; p(\bar{X}/c_k).$$

The probability $p(\bar{X}/c_i)$ is the conditional probability of $\bar{X}$, given that c_i has occurred. $p(c_i/\bar{X})$ is called the likelihood function of class c_i. Substitute the likelihood function into the loss function $r_j(\bar{X})$ and get

$$r_j(\bar{X}) = \sum_{i=1}^{N} l_{ij}\, p(c_i)\, p(\bar{X}/c_i).$$

Note that $p(\bar{X})$ is a common factor in $r_j(\bar{X})$ for all j and is dropped. In general, the pattern $\bar{X}$ is assigned to class c_i if $r_i(\bar{X}) < r_j(\bar{X})$ for $j = 1, N$ and $j \neq i$.

Suppose there is no loss for a correct classification and a unit loss for a misclassification, then $lii = 0$ for all $i = 1, N$ and $lij = 1$ otherwise. Put these values into $r_j(\bar{X})$ and get

$$r_j(\bar{X}) = \sum_{i=1}^{N} p(c_i)\ p(\bar{X}/c_i) - p(c_j)\ p(\bar{X}/c_j)$$

$$r_j(\bar{X}) = p(\bar{X}) - p(c_j)\ p(\bar{X}/c_j).$$

This says to assign the observation $\bar{X}$ to class c_i if

$$p(\bar{X}) - p(c_j)\, p(\bar{X}/c_j) < p(\bar{X}) - p(c_j)\, p(\bar{X}/c_j)$$

for all $j = 1, N$ and $j \neq i$.

This relation is written in its final form as

$$p(c_i)\, p(\bar{X}/c_i) > p(c_j)\, p(\bar{X}/c_j).$$

APPENDIX B.

Bayes Classifier for Multivariate Normal Signatures • From appendix A, the Bayes classifier for the special case of no loss for a correct call and unit loss for a miscall is

$$p(c_i)\, p(\bar{X}/c_i) > p(c_j)\, p(\bar{X}/c_j).$$

This relation says to assign the texture vector $\bar{X}$ to class c_i if the inequality holds for all $j \neq i$. The probability $p(c_k)$ is the prior probability of the k^{th} class, and the probability $p(\bar{X}/c_k)$ is the conditional probability of $\bar{X}$, given that the k^{th} class has occurred.

The classification problem has been reduced to calculating N decision functions $d_i = p(c_i)\, p(\bar{X}/c_i)$ and assigning $\bar{X}$ to the class associated with the largest d_i. Assume that texture signatures are distributed according to the Multivariate Normal distribution

$$p(X/c_i) = \frac{1}{(2\pi)^{N/2}\, |\Sigma_i|^{1/2}}\, e^{-\frac{1}{2}(\bar{X} - \mu_i)^T\, \Sigma_i^{-1}(\bar{X} - \mu_i)}$$

where μ_i and Σ_i are the mean and covariance matrix of the distribution.

The decision functions can be simplified by taking natural logarithms of the functions and using the logarithms as decision functions. This is a valid operation since the log function is a monotonically increasing function.

$$\ln(d_i) = \ln p(c_i) + \ln p(\bar{X}/c_i)$$

$$\ln(d_i) = \ln p(c_i) - \frac{N}{2} \ln 2\pi - \tfrac{1}{2}\ln |\Sigma_i| - \tfrac{1}{2}(\bar{X} - \mu_i)^T \Sigma_i^{-1} (\bar{X} - \mu_i)$$

The value $\frac{N}{2} \ln 2\pi$ does not depend on i and can be ignored. The decision functions turn out to be

$$d_i = \ln p(c_i) - \tfrac{1}{2}\ln |\Sigma_i| - \tfrac{1}{2} (\bar{X} - \mu_i)^T \Sigma_i^{-1} (\bar{X} - \mu_i)$$

Note that the set of decision functions, d_i is composed of constant, linear, and quadratic terms in X, which means that the multivariate normal Baysian classifier places, at most, a second order surface between pairs of texture classes.

APPENDIX C

Divergence • Divergence is a measure of the difficulty of distinguishing between two hypotheses H_1 and H_2.[7] Note that the hypotheses can be regarded as "belonging to Class 1 or belonging to Class 2." Suppose the pattern vector $\bar{X}$ can occur in conjunction with two mutually exclusive events c_1 and c_2. From Bayes' theorem

$$p(c_1/\bar{X}) = \frac{p(c_1) * p(\bar{X}/c_1)}{p(c_1) * p(\bar{X}/c_1) + p(c_2) * p(\bar{X}/c_2)}$$

$$p(c_2/\bar{X}) = \frac{p(c_2) * p(\bar{X}/c_2)}{p(c_1) * p(\bar{X}/c_1) + p(c_2) * p(\bar{X}/c_2)}$$

and

$$\frac{p(c_1/\bar{X})}{p(c_2/\bar{X})} = \frac{p(c_1) * p(\bar{X}/c_1)}{p(c_2) * p(\bar{X}/c_2)}$$

Take the logarithm of both sides of the equality and get

$$\log \frac{p(\bar{X}/c_1)}{p(\bar{X}/c_2)} = \log \frac{p(c_1/\bar{X})}{p(c_2/\bar{X})} - \log \frac{p(c_1)}{p(c_2)}$$

Consider the right side of this last equality.

$\frac{p(c_1/\bar{X})}{p(c_2/\bar{X})}$ is the odds in favor of c_1 over c_2, given that $\bar{X}$ has occurred.

---

[7] Solomon Kullback, *Information Theory and Statistics*, Dover Publications, Inc., New York, 1968.

$\frac{p(c_1)}{p(c_2)}$ is the odds in favor of c_1 over c_2 before the observation $\bar{X}$ is made.

Then the logarithm of the likelihood ratio is defined as the information contained in the observation $\bar{X}$ for discriminating in favor of c_1 over c_2. The average information for discriminating in favor of c_1 over c_2 is

$$I(1:2) = \int p(\bar{X}/c_1) \log \frac{p(\bar{X}/c_1)}{p(\bar{X}/c_2)} \, d\bar{X}.$$

Suppose the $\bar{X}$ is from a multivariate normal distribution and that natural logarithms are used, then

$$I(1:2) = \tfrac{1}{2}\ln \frac{|\Sigma_2|}{|\Sigma_1|} + \tfrac{1}{2}T_r \Sigma_1 (\Sigma_2^{-1} - \Sigma_1^{-1})$$

$$+ \tfrac{1}{2}T_r \Sigma_2^{-1}(\bar{u}_1 - \bar{u}_2)(\bar{u}_1 - \bar{u}_2)^T.$$

The details of this derivation can be reviewed in Chapter 9 of Kullback.[8] Note that T_r pertains to the trace (sum of diagonal elements) of a matrix. There are several worthwhile properties of $I(1:2)$ given in the reference. For example, if $\bar{X}$ and $\bar{Y}$ are independent observations, then the amount of information for discriminating in favor of c_1 over c_2 is $I(1:2;\bar{X},\bar{Y}) = I(1:2;\bar{X}) + I(1:2;\bar{Y})$.

If $I(1:2)$ is the amount of information in the observation $\bar{X}$ for discriminating in favor of Class c_1 over Class c_2, then by a similar argument, $I(2:1)$ is the amount of information in $\bar{X}$ for discriminating in favor of c_2 over c_1.

[8] Solomon Kullback, *Information Theory and Statistics*, Dover Publications, Inc., New York, 1968, Chapter 9.

$$I(2:1) = \int p(\bar{X}/c_2) \log \frac{p(\bar{X}/c_2)}{p(\bar{X}/c_1)} d\bar{X}$$

or

$$I(2:1) = -\int p(\bar{X}/c_2) \log \frac{p(\bar{X}/c_1)}{p(\bar{X}/c_2)} d\bar{X}$$

Divergence is defined to be

$$J(1, 2) = I(1:2) + I(2:1)$$

or

$$J(1, 2) = \int \left(p(\bar{X}/c_1) - p(\bar{X}/c_2)\right) \log \frac{p(\bar{X}/c_1)}{p(\bar{X}/c_2)} d\bar{X}.$$

Divergence is a measure of the difficulty of discriminating between Class c_1 and Class c_2. Note that from symmetry $J(1, 2) = J(2, 1)$. If the observation $\bar{X}$ is from a multivariate normal distribution, then

$$J(1, 2) = \tfrac{1}{2}\mathrm{Tr}(\Sigma_1 - \Sigma_2)(\Sigma_2^{-1} - \Sigma_1^{-1}) + \tfrac{1}{2}\mathrm{Tr}(\Sigma_1^{-1} + \Sigma_2^{-1})(\bar{u}_1 - \bar{u}_2)(\bar{u}_1 - \bar{u}_2)^t.$$

The details of this deviation can be reviewed in Chapter 9 of Kullback.[9]

[9]Solomon Kullback, *Information Theory and Statistics*, Dover Publication,Inc.,New York, 1968,Chapter 9.

The following properties of $J(1, 2)$ are derived there:

- If $\bar{X}$ and $\bar{Y}$ are independent obserations, then $J(1, 2; \bar{X}, \bar{Y}) = J(1, 2; \bar{X}) + J(1, 2; \bar{Y})$.
- $J(1, 2) > 0$.
- The above properties imply that

$$J(1, 2; \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_N, \bar{X}_{N+1}) > J(1, 2; \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_N)$$

The last property says that new observations can be evaluated for their discriminatory power.

Large values of $J(1, 2)$ indicate more power in discriminating between two classes. This can be seen in the following example. Suppose $\Sigma_1 = \Sigma_2 = \Sigma$ and that a measure of the discriminatory power of an observation for distinguishing between $\bar{u}_1$ and $\bar{u}_2$ is desired. In this case,

$$J(1, 2; \bar{u}) = (\bar{u}_1 - \bar{u}_2)\, \Sigma^{-1} (\bar{u}_1 - \bar{u}_2)^T$$

and in the case of univariates,

$$J(1, 2; u) = \frac{(u_1 - u_2)^2}{\sigma^2}$$

If the population means are nearly equal, then there is little discriminatory power in one observation. If the means are very different, then there is more discriminatory power in a single observation, and $J(1, 2; u)$ is also larger.

APPENDIX D.

Confusion Matrices • The classification and relaxation results from the numerical experiment are listed here. The $(ij)^{th}$ entry in each table refers to the percentage of times the i^{th} class was classified as the j^{th} class. For example, consider the two (6 x 6) matrices of table D1 that pertain to window size 15. The first matrix M pertains to initial results from maximum likelihood, and the second matrix R pertains to relaxation results. The entry $m_{33} = 61.5$ indicates that 61.5 percent of class 3 was called correctly, whereas $m_{36} = 10.6$ says that 10.6 percent of class 3 was called class 6. Similarly, $m_{66} = 60.2$ says that 60.2 percent of class 6 was called correctly, whereas $m_{63} = 3.7$ says that 3.7 percent of class 6 was called class 3. There was a total misclassification error between these two classes of 14.3 percent. The modified results after relaxation can be determined by reviewing the corresponding elements of the R matrix.

TABLE D1. Confusion Matrices For Scene A. PANC

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
	Class	1	2	3	4	5	6	1	2	3	4	5	6
21	1	87.7	3.7	1.2	0.0	4.9	2.5	97.5	2.5	0.0	0.0	0.0	0.0
	2	0.0	91.7	1.4	0.0	2.8	4.2	0.0	100.0	0.0	0.0	0.0	0.0
	3	15.4	0.0	72.1	0.0	4.8	7.7	0.0	0.0	100.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	1.6	0.0	0.0	0.0	96.8	1.6	0.0	0.0	0.0	0.0	100.0	0.0
	6	0.0	0.0	.9	3.7	.9	94.4	0.0	0.0	0.0	0.0	0.0	100.0
15	1	67.9	6.2	2.5	7.4	9.9	6.2	97.5	2.5	0.0	0.0	0.0	0.0
	2	0.0	88.9	0.0	1.4	9.7	1.4	0.0	100.0	0.0	0.0	0.0	0.0
	3	4.8	1.0	51.5	12.5	9.6	10.6	0.0	0.0	98.1	0.0	1.9	0.0
	4	0.0	1.2	0.0	95.2	2.4	1.2	0.0	0.0	0.0	100.0	0.0	0.0
	5	0.0	7.9	0.0	1.6	82.5	7.9	0.0	0.0	0.0	0.0	100.0	0.0
	6	1.9	3.6	3.7	20.4	8.3	60.2	0.0	0.0	.9	0.0	.9	98.1
9	1	82.7	0.0	12.3	1.2	3.0	3.7	100.0	0.0	0.0	0.0	0.0	0.0
	2	23.6	15.3	30.6	8.3	1.4	20.3	11.1	2.8	69.4	0.0	0.0	16.7
	3	18.6	0.0	71.2	9.6	1.0	7.7	2.9	0.0	97.1	0.0	0.0	0.0
	4	3.6	0.0	9.5	76.2	1.2	9.5	1.2	0.0	2.4	92.9	0.0	3.6
	5	12.7	0.0	44.4	4.8	19.0	19.0	6.3	0.0	79.4	0.0	1.6	12.7
	6	11.1	.9	21.3	4.6	3.0	62.0	0.0	0.0	2.8	0.0	0.0	97.2

TABLE D2. Confusion Matrices For Scene A. IR

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
	Class	1	2	3	4	5	6	1	2	3	4	5	6
21	1	82.7	2.5	0.0	12.3	2.5	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	2.8	94.4	0.0	0.0	2.8	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	3.8	2.9	81.7	1.9	1.0	8.7	0.0	0.0	100.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	0.0	3.2	1.6	0.0	95.2	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	6	.9	.9	.9	1.9	0.0	95.4	0.0	0.0	0.0	0.0	0.0	100.0
15	1	67.9	3.7	2.5	22.2	2.5	1.2	100.0	0.0	0.0	0.0	0.0	0.0
	2	2.8	79.2	4.2	1.4	12.5	0.0	1.4	98.6	0.0	0.0	0.0	0.0
	3	1.9	3.8	67.9	10.6	3.8	12.5	0.0	0.0	100.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	98.8	0.0	1.2	0.0	0.0	0.0	100.0	0.0	0.0
	5	0.0	0.0	1.6	1.6	96.8	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	6	.9	2.8	3.7	9.3	.9	82.4	0.0	0.0	0.0	0.0	0.0	100.0
9	1	75.3	6.2	3.7	3.7	6.2	4.9	100.0	0.0	0.0	0.0	0.0	0.0
	2	5.6	79.2	5.6	4.2	2.8	2.8	2.8	97.2	0.0	0.0	0.0	0.0
	3	14.6	7.7	64.4	1.9	7.7	7.7	0.0	0.0	98.0	0.0	0.0	1.0
	4	19.0	3.6	2.4	61.9	1.2	11.9	4.8	0.0	0.0	95.2	0.0	0.0
	5	7.9	14.3	9.5	1.6	61.9	4.8	1.6	3.2	1.6	0.0	92.1	1.6
	6	3.7	11.1	10.2	3.7	4.6	66.7	0.0	2.8	.9	0.0	0.0	96.3

TABLE D3. Confusion Matrices For Scene B. PANC

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
	Class	1	2	3	4	5	6	1	2	3	4	5	6
21	1	98.3	1.7	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	11.1	80.8	1.0	1.0	4.0	2.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	2.5	4.9	79.0	0.0	9.9	3.7	0.0	0.0	100.0	0.0	0.0	0.0
	4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	3.5	.7	2.8	4.2	86.7	2.1	2.8	0.0	0.0	0.0	96.5	.7
	6	11.1	1.2	1.2	0.0	3.7	82.7	0.0	0.0	0.0	0.0	0.0	100.0
15	1	96.7	1.7	.8	0.0	.8	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	29.3	58.6	3.0	2.0	4.0	3.0	1.0	98.0	0.0	0.0	1.0	0.0
	3	0.2	2.5	79.0	3.7	3.7	4.9	1.2	1.2	97.5	0.0	0.0	0.0
	4	0.0	0.0	0.0	94.8	5.2	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	3.5	1.4	2.8	14.7	75.5	2.1	4.9	0.0	0.0	2.1	93.0	0.0
	6	22.2	4.9	1.2	1.2	6.2	64.2	3.7	0.0	0.0	0.0	0.0	96.3
9	1	76.0	5.0	2.5	7.4	5.8	3.3	98.3	0.0	0.0	.8	.8	0.0
	2	17.2	56.6	5.1	6.1	10.1	5.1	1.0	99.0	0.0	0.0	0.0	0.0
	3	8.6	6.2	48.1	16.0	19.8	1.2	1.2	0.0	97.5	0.0	0.0	1.2
	4	1.3	0.0	1.3	83.1	13.0	1.3	0.0	0.0	0.0	100.0	0.0	0.0
	5	8.4	0.0	3.5	22.4	62.2	3.5	2.8	0.0	0.0	5.5	91.6	0.0
	6	12.3	12.3	0.0	9.9	8.6	56.8	1.2	0.0	0.0	1.2	1.2	96.3

TABLE D4. Confusion Matrices For Scene B. IR

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
		1	2	3	4	5	6	1	2	3	4	5	6
21	1	95.6	4.1	0.[illegible]	0.0	0.0	.8	100.0	0.0	0.0	0.0	0.0	0.0
	2	2.0	90.9	2.0	2.0	0.0	3.0	0.0	100.0	[illegible]	0.0	0.0	0.0
	3	0.0	4.9	76.5	7.4	8.6	2.5	0.0	0.0	97.5	2.5	0.0	0.0
	4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	3.5	2.1	4.9	4.2	85.3	0.0	2.8	1.4	0.0	2.8	93.0	0.0
	6	7.4	2.5	0.0	0.0	0.0	90.1	0.0	0.0	0.0	0.0	0.0	100.0
15	1	81.8	5.8	1.7	.8	0.0	9.9	100.0	0.0	0.0	0.0	0.0	0.0
	2	13.1	67.7	1.0	4.0	5.1	9.1	0.0	99.0	0.0	1.0	0.0	0.0
	3	6.2	9.9	55.6	13.6	11.1	3.7	0.0	2.5	91.4	6.2	0.0	0.0
	4	1.3	0.0	0.0	97.4	1.3	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	4.9	0.0	4.2	7.0	82.5	1.4	4.2	0.0	0.0	3.5	91.6	.7
	6	7.4	3.7	0.0	1.2	0.0	87.7	0.0	0.0	0.0	0.0	0.0	100.0
9	1	62.8	11.6	.8	5.0	13.2	6.6	96.7	.8	0.0	1.7	.8	0.0
	2	8.1	65.7	1.0	5.1	14.1	6.1	0.0	98.0	0.0	0.0	2.0	0.8
	3	7.4	7.4	19.8	11.1	40.9	7.4	1.2	3.7	19.8	4.9	67.9	2.5
	4	3.9	0.0	0.0	74.0	22.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	4.9	1.4	.7	5.6	85.3	2.1	.7	1.4	0.0	2.1	95.8	0.0
	6	3.7	9.9	0.0	4.9	3.7	77.8	0.0	0.0	0.0	0.0	0.0	100.0

TABLE D5. Confusion Matrices For Scene C. PANC

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS					RELAXATION RESULTS				
	Class	1	2	3	4	5	1	2	3	4	5
21	1	96.6	3.4	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	2	8.5	91.5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	3	1.9	1.9	93.3	1.9	1.0	0.0	0.0	100.0	0.0	0.0
	4	0.0	0.0	2.4	97.6	0.0	0.0	0.0	0.0	100.0	0.0
	5	0.0	0.0	0.0	8.7	91.3	0.0	0.0	0.0	0.0	100.0
15	1	95.7	3.4	0.0	.9	0.0	100.0	0.0	0.0	0.0	0.0
	2	20.6	77.6	1.8	0.0	0.0	0.0	100.0	0.0	0.0	0.0
	3	7.7	12.5	76.9	0.0	2.9	0.0	0.0	100.0	0.0	0.0
	4	0.0	1.2	6.0	88.1	4.8	0.0	0.0	0.0	100.0	0.0
	5	0.0	0.0	2.9	7.7	89.4	0.0	0.0	1.0	0.0	99.0
9	1	92.3	4.3	.9	0.0	2.6	100.0	0.0	0.0	0.0	0.0
	2	32.1	58.2	1.8	3.0	4.8	3.6	90.4	0.0	0.0	0.0
	3	24.0	12.5	45.2	6.7	11.5	8.7	1.9	89.4	0.0	0.0
	4	4.8	8.3	1.2	78.6	7.1	0.0	2.4	0.0	97.6	0.0
	5	0.0	2.9	4.8	7.7	84.6	0.0	1.0	0.0	1.0	99.1

TABLE D6. Confusion Matrices For Scene C. IR

WINDOW SIZE	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
	Class \ Class	1	2	3	4	5	Class	1	2	3	4	5
21	1	98.3	1.7	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0
	2	10.3	89.7	0.0	0.0	0.0	2	0.0	100.0	0.0	0.0	0.0
	3	1.0	1.0	98.1	0.0	0.0	3	0.0	0.0	100.0	0.0	0.0
	4	0.0	0.0	6.0	92.9	1.2	4	0.0	0.0	0.0	100.0	0.0
	5	0.0	0.0	3.8	1.0	95.2	5	0.0	0.0	0.0	0.0	100.0
15	1	88.0	12.0	0.0	0.0	0.0	1	100.0	0.0	0.0	0.0	0.0
	2	7.9	89.7	2.4	0.0	0.0	2	.6	99.4	0.0	0.0	0.0
	3	1.9	2.9	92.3	1.9	1.0	3	0.0	0.0	100.0	0.0	0.0
	4	1.2	1.2	4.8	89.3	3.6	4	0.0	0.0	0.0	100.0	0.0
	5	0.0	0.0	8.7	2.9	88.5	5	0.0	0.0	0.0	0.0	100.0
9	1	76.9	16.2	1.7	3.4	1.7	1	100.0	0.0	0.0	0.0	0.0
	2	9.7	81.8	.6	4.8	3.0	2	0.0	99.4	0.0	0.0	.6
	3	5.8	9.6	42.3	26.9	15.4	3	0.0	0.0	95.2	3.8	1.0
	4	0.0	3.0	4.8	85.7	6.0	4	0.0	0.0	0.0	100.0	0.0
	5	0.0	1.9	5.8	14.4	77.9	5	0.0	0.0	0.0	0.0	100.0

TABLE D7. Confusion Matrices For Scene E. PANC

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
		1	2	3	4	5	6	1	2	3	4	5	6
21	1	95.1	3.7	0.0	1.2	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	5.1	94.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	88.2	2.6	0.0	9.2	0.0	0.0	100.0	0.0	0.0	0.0
	4	0.0	0.0	1.2	97.5	1.2	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0
15	1	96.3	2.5	1.2	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	7.1	88.9	1.0	3.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	.7	.7	80.4	2.0	.7	15.7	0.0	0.0	100.0	0.0	0.0	0.0
	4	2.5	0.0	9.9	80.2	0.0	7.4	0.0	0.0	0.0	98.8	0.0	1.2
	5	0.0	0.0	3.3	0.0	90.0	6.7	0.0	0.0	0.0	0.0	100.0	0.0
	6	0.0	0.0	3.7	0.0	0.0	96.3	0.0	0.0	0.0	0.0	0.0	100.0
9	1	49.4	43.2	2.5	4.9	0.0	0.0	87.7	12.3	0.0	0.0	0.0	0.0
	2	0.0	98.0	1.0	1.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	2.0	14.4	64.7	2.6	.7	15.7	0.0	.7	98.7	0.0	0.0	.7
	4	4.9	24.7	11.1	45.7	0.0	13.6	0.0	2.5	0.0	97.5	0.0	0.0
	5	0.0	6.7	3.3	0.0	90.0	0.0	0.0	3.3	0.0	0.0	96.7	0.0
	6	1.2	8.6	6.2	0.0	0.0	84.0	0.0	0.0	0.0	0.0	0.0	100.0

TABLE D8. Confusion Matrices For Scene E. IR

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS						RELAXATION RESULTS					
		1	2	3	4	5	6	1	2	3	4	5	6
21	1	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	96.1	1.3	0.0	2.6	0.0	0.0	100.0	0.0	0.0	0.0
	4	1.2	0.0	7.4	87.7	0.0	3.7	0.0	0.0	0.0	100.0	0.0	0.0
	5	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	6	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0
15	1	96.3	3.7	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	3.0	96.0	0.0	1.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0
	3	0.0	0.0	83.7	2.0	.7	13.7	0.0	0.0	100.0	0.0	0.0	0.0
	4	3.7	0.0	9.9	72.8	0.0	13.6	0.0	0.0	0.0	100.0	0.0	0.0
	5	0.0	0.0	10.0	0.0	90.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
	6	0.0	0.0	8.6	0.0	0.0	91.4	0.0	0.0	0.0	0.0	0.0	100.0
9	1	90.1	4.9	0.0	4.9	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0
	2	11.1	86.9	0.0	1.0	1.0	0.0	1.0	99.0	0.0	0.0	0.0	0.0
	3	9.2	1.3	52.3	30.1	1.3	5.9	0.0	0.0	80.4	19.6	0.0	0.0
	4	11.1	3.7	1.2	84.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
	5	6.7	6.7	6.7	0.0	80.0	0.0	0.0	3.3	0.0	0.0	96.7	0.0
	6	4.9	1.2	8.6	32.1	0.0	53.1	0.0	0.0	0.0	9.9	0.0	90.1

TABLE D9. Confusion Matrices For Scene H. PANC

WINDOW SIZE	Class	MAXIMUM LIKELIHOOD RESULTS				Class	RELAXATION RESULTS			
		1	2	3	4		1	2	3	4
21	1	85.1	5.0	7.4	2.5	1	100.0	0.0	0.0	0.0
	2	3.4	92.3	4.8	0.0	2	0.0	100.0	0.0	0.0
	3	3.3	7.2	81.0	8.5	3	0.0	0.0	100.0	0.0
	4	0.0	3.3	0.0	96.7	4	0.0	0.0	0.0	100.0
15	1	68.6	14.9	14.9	1.7	1	100.0	0.0	0.0	0.0
	2	7.1	86.3	0.0	.6	2	0.0	100.0	0.0	0.0
	3	3.9	9.8	81.0	5.2	3	0.0	0.0	100.0	0.0
	4	2.2	4.4	3.3	90.0	4	0.0	0.0	0.0	100.0
9	1	57.6	28.1	5.8	2.5	1	99.2	.8	0.0	0.0
	2	8.3	82.1	6.5	3.0	2	0.0	100.0	0.0	0.0
	3	16.3	30.7	43.6	9.2	3	1.3	31.4	67.3	0.0
	4	8.9	14.4	2.2	74.4	4	0.0	0.0	0.0	100.0

TABLE D10. Confusion Matrices For Scene H. IR

WINDOW SIZE	Class / Class	MAXIMUM LIKELIHOOD RESULTS 1	2	3	4	Class / Class	RELAXATION RESULTS 1	2	3	4
21	1	86.8	4.1	7.4	1.7	1	100.0	0.0	0.0	0.0
	2	2.4	87.5	5.4	4.8	2	0.0	100.0	0.0	0.0
	3	5.9	1.3	88.9	3.9	3	0.0	0.0	100.0	0.0
	4	1.1	3.3	0.0	95.6	4	0.0	0.0	0.0	100.0
15	1	81.8	11.6	5.8	.8	1	100.0	0.0	0.0	0.0
	2	2.4	84.5	9.5	3.6	2	0.0	100.0	0.0	0.0
	3	14.4	8.5	74.5	2.6	3	0.0	0.0	100.0	0.0
	4	1.1	6.7	5.6	86.7	4	0.0	0.0	0.0	100.0
9	1	55.4	8.3	30.6	5.8	1	97.5	0.0	2.5	0.0
	2	7.7	86.1	14.9	11.3	2	0.0	96.4	1.8	1.8
	3	7.8	9.2	69.3	13.7	3	0.0	0.0	100.0	0.0
	4	2.2	3.3	8.9	85.6	4	0.0	0.0	0.0	100.0